THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 277, No. 22, Issue of May 31, pp. 19594–19599, 2002

Printed in U.S.A.

A Novel Intragenic Sequence Enhances Initiator-dependent Transcription in Human Embryonic Kidney 293 Cells*

Received for publication, February 5, 2002, and in revised form, March 25, 2002
Published, JBC Papers in Press, March 26, 2002, DOI 10.1074/jbc.M201193200

Chiara Abrescia‡, Eliana De Gregorio‡, Mattia Frontini§, Roberto Mantovani§, and Pierpaolo Di Nocera‡¶

From the ‡Dipartimento di Biologia e Patologia Cellulare e Molecolare, Università degli Studi di Napoli Federico II, Via S. Pansini 5, 80131 Napoli, and the §Dipartimento di Biologia Animale, Università di Modena e Reggio, Via Campi 213/d, 41100 Modena, Italy

In a variety of *Drosophila* TATA-less promoters, transcription is directed by initiator (Inr) sequences, which are faithfully and efficiently recognized only when flanked 3′ by the downstream promoter element (DPE). This motif, which is conserved at ~30 bp from the RNA start site, is viewed as a downstream counterpart to the TATA box, and is recognized by the general transcription factor (TF) IID. By transient expression assays in human embryonic kidney 293 cells, we show that DE1 (distal element 1), a DNA motif located at residues +23 to +29, sustains faithful Inr-dependent transcription as efficiently as the DPE. Transcription significantly increased when DE1 and DPE sequences were adjacently placed on the same template. Results emerging from *in vivo* RNA analyses matched electrophoretic mobility shift assay data. In agarose-electrophoretic mobility shift assays, retarded DNA-protein complexes resulting from the interaction of human holo-TFIID with either Inr^+/DPE^+ or $Inr^+/DE1^+$ promoters were formed at comparable levels, whereas binding of TFIID to both DE1 and DPE motifs was 2-fold increased. The strict requirement for spacing between the Inr and DPE was not observed for DE1, as locating the motif 4 bp away from the +1 site did not impair transcriptional enhancement. DE1 sequences may be common to many promoters and be overlooked because of their poor sequence homology.

A key step in the formation of functional transcription initiation complexes is the recognition of promoter sequences by components of the general transcription machinery. The core promoter sequence context has a significant influence on both the overall efficiency of gene transcription and the ability of individual genes to respond to transcription activators (1). In pol II[1]-dependent transcriptional units, distinct DNA elements have been found to be involved in core promoter function. The TATA box, a sequence located 25–30 bp upstream of the RNA start site, is the key positioning DNA element in many pol II genes (1, 2). The TATA box is recognized by the TATA-binding protein (TBP) subunit of the TFIID complex, a general pol II transcription factor endowed with the ability to recognize promoter DNA (2, 3). In some promoters, the TATA box is immediately preceded by the TFIIB recognition element, fitting the consensus 5′-(G/C)-(G/C)-(G/A)-C-G-C-C-3′. The transcription factor TFIIB plays a central role in preinitiation complex assembly, providing a bridge between promoter-bound TFIID and RNA polymerase II, and TFIIB recognition element increases the affinity of TFIIB for the promoter (4). In many promoters, the TATA element is missing, and is functionally replaced by the initiator (Inr), a stretch of 5–7 residues spanning the RNA start site (5, 6). The Inr is also recognized by TFIID, but physical interactions are mediated by some of the TBP-associated factors, or TAF_{II}s (7, 8). TATA and Inr are functionally exchangeable modules and may coexist in the same gene. The core promoter structure found in a given gene may reflect a preference of the regulators of that gene, and some activators stimulate preferentially TATA-containing or Inr-containing core promoters (9–11).

An additional core promoter module is the downstream promoter element, or DPE. This sequence, conserved ~30 bp downstream from the RNA start site in a variety of *Drosophila* TATA-less promoters, greatly enhances the activity of upstream Inr modules (12–17). DPE interacts with specific components of the *Drosophila* TFIID complex (dTAFII40 and dTAFII60; see Ref. 15). Regions downstream of transcriptional start sites recognized by TFIID, but exhibiting no sequence similarity to DPE, have been identified in a variety of promoters (18–22).

Little is known about the role that intragenic sequences have in promoter recognition and activation in human cells. In this work, we analyzed the transient expression profile of constructs in which Inr sequences are flanked by different types of downstream promoter sequences in human embryonic kidney (HEK 293) cells. Inr-dependent transcription is enhanced by a core DPE sequence located at residues +30/+33. The same holds true for a DNA element called DE1 located at residues +23/+29. DPE and DE1 modules synergize in stimulating transcription *in vivo* and are independently capable, as revealed by agarose-EMSA, to interact with human holo-TFIID.

MATERIALS AND METHODS

Construction of Plasmids—Plasmids described in this work carrying artifical promoters are all derivatives of p8GAL4, a modified pEMBL8CAT vector in which a 54-bp module containing one binding site for the transcriptional activator Gal4 had been inserted at the *Bam*HI site. Sequences homologous to the Inr regions of the *Drosophila* Doc and I LINE promoters (16) have been inserted between the *Bam*HI

* This work was supported by grants from Ministero dell'Università e Ricerca Scientifica, Project "Dinamica della Cromatina nella Espressione Genica" (to P. P. D. N. and R. M.), and Associazione Italiana della Ricerca sul Cancro (to R. M.). The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

¶ To whom correspondence should be addressed. Tel.: 39-81-7462059; Fax: 39-81-7703285; E-mail: dinocera@cds.unina.it.

[1] The abbreviations used are: pol II, RNA polymerase II; CAT, chloramphenicol acetyltransferase; DCE, downstream core element; DE1, distal element 1; DPE, downstream promoter element; EMSA, electrophoretic mobility shift assay; HEK, human embryonic kidney; hsp70, heat shock protein 70; Inr, initiator; IRF-1, interferon regulatory factor 1; LINE, long interspersed nuclear element; RSV, Rous sarcoma virus; TAF_{II}, TATA-binding protein-associated factor; TBP, TATA-binding protein; TF, transcription factor.

and *Sal*I sites of p8GAL4 to obtain the plasmids G3 and G1, respectively. Constructs G1M, G1X, G1K, G3M, G3X, and G3K, which carry downstream promoter modules, have been obtained by inserting between the *Sal*I and *Hin*dIII sites of either G1 or G3 suitable pairs of complementary oligonucleotides. Plasmids G4M and G5M are derivatives of G3M in which the interval *Bam*HI/*Sal*I spanning the Inr region had been modified. Similarly, the mutant constructs analyzed in Fig. 4 are derivatives of G3M in which the downstream promoter region had been replaced by oligonucleotide pairs having *Sal*I- and *Hin*dIII-compatible termini. To obtain derivatives in which the distance between Inr and downstream sequences increased 4 bp, plasmids of interest were digested with *Sal*I, and reaction products were treated with the Klenow enzyme to fill in gaps prior to ligation and transformation. The control plasmid RSVdel-CAT was obtained by cloning the *Apa*L-*Mlu*I fragment spanning the RSV promoter region in pRSVCAT into the *Hin*dIII site of pEMBL8CAT, and subsequently removing the *Hin*dIII-*Nco*I fragment including most of the CAT coding region. The GAL4-Sp1 plasmid encodes a chimeric Sp1 GAL4 protein containing residues 50–161 of the human Sp1 protein. In all cloning procedures, incompatible termini were blunt-ended by T4 polymerase before ligation. The sequences of the promoter regions analyzed were confirmed by nucleotide sequence analysis.

Cell Culture and DNA Transfections—HEK 293 cells were grown in Dulbecco's modified Eagle's medium supplemented with 10% fetal bovine serum. Transfection experiments were performed with the standard calcium-phosphate method. Approximately 6×10^5 cells, seeded at a density of 1.2×10^5 cells/ml 24 h prior to transfection, were cotransfected with 10 μg of the plasmid of interest, 0.1 μg of GAL4-Sp1 plasmid, 0.2 μg of RSVdel-CAT plasmid for control of transfection efficiency. Cells were recovered 48 h after transfection, and the activity of constructs was assayed at the RNA level.

RNA Analyses—Total RNA was isolated by using the acid guanidinium thiocyanate/phenol/chloroform single-step extraction method (23). Primer extension assay experiments were performed essentially as described (16). Reaction products were resolved on 8% (w/v) polyacrylamide, 8 M urea gels. Co-electrophoresed sequencing ladders were generated by the dideoxy chain termination method utilizing double-stranded DNA templates. The CAT primer used both to obtain sequencing ladders and to detect transcripts directed by the different promoters constructs has been described previously (24). Transcripts driven by the RSV promoter in the RSVdel-CAT plasmid were detected by using the NCO primer (5′-AGCGGCATCAGCACCTTGTCGCCTTGCGTA-3′), a synthetic 30-mer complementary to a distal interval of the CAT gene sense strand.

Purification of Holo-TFIID and EMSA Analysis—Holo-TFIID was immunopurified from HeLa cells with an anti-TBP antibody as previously detailed (25, 26). EMSAs of TFIID in agarose gels were performed as described in Ref. 27. Three independent preparations of purified TFIID were used in EMSAs. The GAL4-NF-YA fusion protein (28) was obtained by an *in vitro* transcription-translation coupled reticulocyte lysate system (Promega). One μl of GAL4-NF-YA-containing extracts and 10,000 cpm of ^{32}P-labeled fragments were mixed in 10 μl of NF-Y buffer (20 mM Hepes, pH 7.9, 50 mM NaCl, 5% glycerol, 5 mM $MgCl_2$, 1 mM dithiothreitol) and incubated for 20 min at 30 °C. Samples were loaded onto 4.5% polyacrylamide gels (acrylamide/bisacrylamide, 29:1) and electrophoresed in 0.5× TBE buffer. Gels were run at 150 V for 60 min, transferred on no. 3MM paper, and exposed. PCR fragments tested in EMSA analysis span residues −66 to +61 of all promoters but G3, in which the amplified region spans residues −66 to +55. Amplification was obtained by using two synthetic oligomers, the ^{32}P-5′-end-labeled CAT II 30-mer (5′-TCCTTAGCTCCTGAAAATCTCGCCAAGCTT-3′), complementary to the pEMBL8CAT sense strand, and the GT1 54-mer (5′-TCTCGAGCTGCAGCGGAGACTGTCCTCCGAGATCTCTATCACTGATAGGGATCG-3′), homologous to the −66/−12 interval spanning the binding site for the transcriptional activator Gal4 included in each promoter.

RESULTS

Activation of Inr-dependent Transcription in HEK 293 Cells—In several *Drosophila* TATA-less promoters, the DPE promoter element fits the consensus AG(A/T)CGTGY (12, 14). Statistical and biochemical analyses indicate that the 4-bp core DPE sequence G(A/T)CG is sufficient to stimulate Inr transcription (15, 17). In some *Drosophila* long interspersed nuclear element (LINE) promoters, transcription is regulated by complex intragenic regions including DPEs and additional DNA

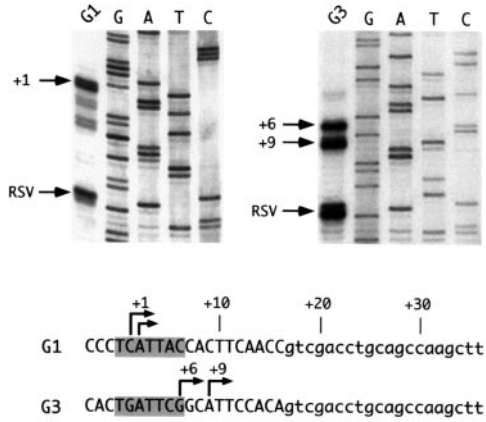

FIG. 1. **Analysis of G1 and G3 promoters.** Total RNA (30 μg) from HEK 293 cells transiently cotransfected with 0.2 μg of the internal reference template RSVdel-CAT plasmid, 0.1 μg of GAL4-Sp1 plasmid, and 10 μg of either G1 (*left panel*) or G3 (*right panel*) plasmid was analyzed by primer extension. Distinct ^{32}P-5′-end-labeled oligomers were used to detect G1-G3 (CAT primer) and RSV (Nco primer) transcripts. Sequencing ladders of the plasmids G1 (*left panel*) and G3 (*right panel*) were obtained by the dideoxy chain termination method using as primer the CAT oligomer. Bands corresponding to major transcripts are marked by *arrows*. The promoter regions of the G1 and G3 templates are aligned at the *bottom*. Inrs (residues −1 to +6) are *highlighted*. *Numbers* are relative to RNA start sites (+1 sites) mapped in *Drosophila* Schneider II cells (16). Vector sequences are in *lowercase letters*.

sequence elements (16). To verify whether sequences flanking DPE in LINE promoters could stimulate transcription in mammalian cells, transient expression assays were carried out in HEK 293 cells. In the base plasmids G1 and G3, promoter DNA is uniquely represented by Inr sequences (see Fig. 1). In the other plasmids, Inr sequences are flanked by ~20-bp-long DNA segments containing, at the correct distance, either DE1 or DPE, or both sequences (see Fig. 2). The DE1 sequence is found immediately upstream of a core DPE motif in the *Drosophila* I promoter (16); DPE corresponds to the 4-bp core DPE sequence GACG found in a variety of *Drosophila* promoters (17). In all templates, a GAL4 recognition sequence is centered ~30 bp upstream of the Inr region. Because Sp1 effectively activates Inr^+ promoters (11), each construct was cotransfected with a plasmid encoding the GAL4-Sp1 activator (28). The plasmid RSVdel-CAT was also cotransfected along each construct to provide an internal control.

Correctly initiated transcripts accounted for most of the signal detected with the G1 construct. However, multiple bands marking the accumulation of minor RNA species initiating within the +4 to +8 interval were also detected. By contrast, faithful +1 transcripts driven by the G3 template accumulated at nearly undetectable levels in HEK 293 cells, the prominent signal obtained corresponding to RNAs initiated at residues +6 and +9 (Fig. 1). This peculiar transcriptional pattern plausibly reflects the activity of a secondary Inr module spanning residues +6 to +13 (GGCATTCC; see Fig. 1). In *Drosophila* Schneider II cells, alternative initiation from this secondary Inr is predominant over initiation from the Inr CTGATTC spanning residues −2 to +6 in the absence of DPE (29). The profile of expression of the G3 Inr dramatically changed upon addition of downstream promoter sequences. The presence of either DE1 (G3K construct), DPE (G3X construct), or both (G3M construct) allowed the detection initiation from residue +1 (Fig. 2). Elongation products were quantitated by PhosphorImager analyses, and the efficiency of faithful transcription initiation evaluated as the ratio of transcripts initiated at residues +1 and +6. Quantitative estimates revealed that the $DE1^+DPE^+$ G3M construct directed faithful transcription initiation ~5- and ~10-fold more efficiently than the $DE1^-DPE^+$ G3X and the

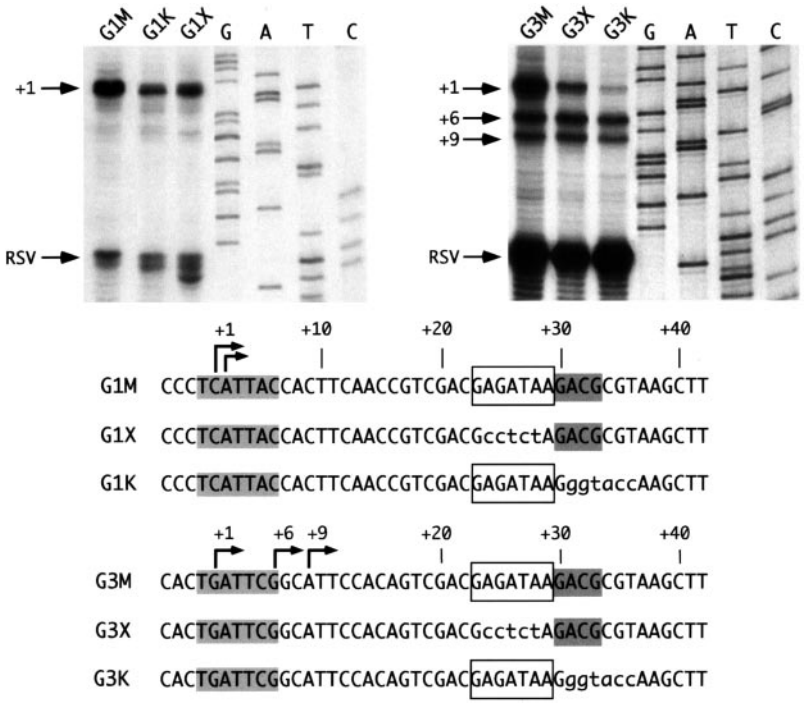

FIG. 2. **Transcriptional reprogramming by DPE and DE1 sequences.** Total RNA (30 μg) from HEK 293 cells transiently cotransfected with 0.2 μg of RSVdel-CAT plasmid, 0.1 μg of GAL4-Sp1 plasmid, and 10 μg of the DNAs indicated at the *top* was analyzed by primer extension as in Fig. 1. Co-electrophoresed sequencing ladders of G1M (*left panel*) and G3M (*right panel*) were obtained by using as primer the CAT oligomer. The promoter regions of the templates assayed are aligned at the *bottom*. Inr, DE1, and DPE sequences are *highlighted*. Base changes altering residues found in G1M and G3M are in *lowercase letters*.

DE1$^+$ DPE$^-$ G3K constructs, respectively.

Thus, DE1 and DPE motifs can reprogram, at comparable levels, the pattern of transcription initiation of the G3 Inr. Moreover, the two motifs synergize in enhancing functional recognition of the +1 site. Similar results were obtained by the analysis of the G1 derivatives G1M, G1K, and G1X (Fig. 2). Comparisons of the autoradiograms shown in Figs. 1 and 2 reveal that initiation at minor sites detected with the parental G1 template was largely inhibited in the three G1 derivatives, each template directing predominantly, if not exclusively, the synthesis of +1 transcripts. To evaluate relative template efficiencies, transcripts directed by the G1 Inr were normalized to transcripts directed by the reference RSVdel-CAT construct. The DE1$^+$DPE$^+$G1M template directed faithful transcription initiation ~2- and ~4-fold more efficiently than the DPE$^+$ G1X and the DE1$^+$ G1K templates, respectively. Differences in the degree of stimulation by the same sequence elements in the G1 and G3 derivatives correlate to differences in the strength of the G1 and G3 Inrs. The G1 Inr, which better fits the optimal Inr consensus (6) and efficiently directs faithful transcription initiation as single module (Fig. 1) is less sensitive to enhancement by downstream promoter sequences. For this reason, we measured the fidelity of initiation as the ratio of transcripts originating from the same template, and subsequent analyses were all carried out with the G3 Inr.

In G3M, both the Inr and the downstream sequences are required to measure efficient transcriptional levels, as shown from the analysis of G4M and G5M, two derivatives in which the *Inr* was variously mutated (Fig. 3). In HEK 293 cells transfected with either plasmid, transcripts initiated at residues +6 and +9 account for most of the detectable signal. Interestingly, transcription initiation at or near residue +1 could still be measured, albeit at very low levels (see bands marked by *gray arrows* in Fig. 3). Heterogeneity in the start sites around position +1 between G4M and G5M may correlate to alternative base selection dictated by the different DNA contexts replacing genuine Inr sequences. These data indicate that, albeit inefficiently, downstream promoter elements contribute to correctly position the transcriptional apparatus even in the absence of a functional Inr.

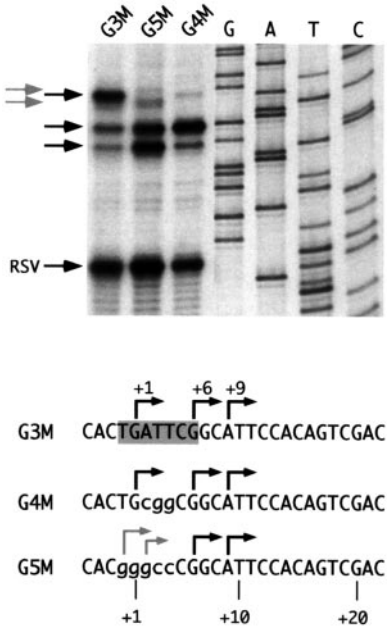

FIG. 3. **Analysis of Inr$^-$ templates.** The template proficiency of G3M, G5M, and G4M in HEK 293 cells was analyzed by primer extension as described in Figs. 1 and 2. Samples were electrophoresed along with a G3M sequence ladder obtained as in Fig. 2. *Black arrows* mark major reaction products, and *gray arrows* minor transcripts driven by the G5M promoter. The G3M, G4M, and G5M start regions are aligned at the *bottom*. Base changes altering the Inr are in *lowercase letters*.

Functional Dissection of the DE1-DPE Region—To characterize the functional interplay between DE1 and DPE, we next transfected HEK 293 cells with derivatives of G3M, in which base changes were introduced within the +23/+33 interval to selectively alter either DNA motif. The base changes and transcriptional proficiencies of the constructs analyzed are reported in Fig. 4. The levels of correct transcription initiation directed by each construct were evaluated by calculating, as in Fig. 2, the +1 *versus* +6 transcript ratio. By looking at templates carrying base changes in the DE1 sequence, mutating positions +25 and +26 (plasmids 34B and 34C) lowered ~2-fold the

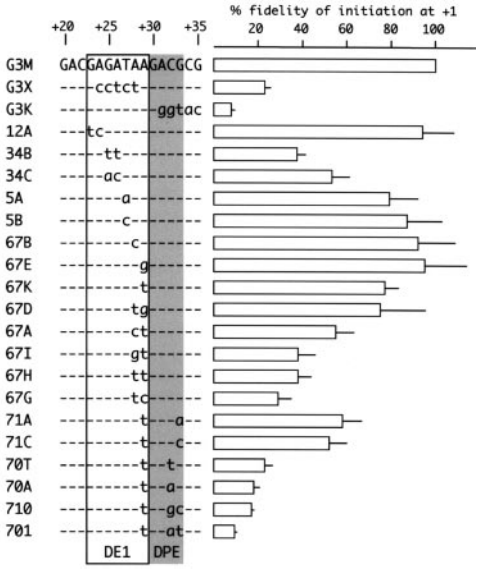

FIG. 4. **Base changes within the downstream promoter region of G3M and transcriptional proficiency.** The downstream promoter regions (residues +20/+35) of the constructs analyzed are aligned to the parental G3M sequence. *Dashes* denote sequence identities. DE1 and DPE motifs are *highlighted*. *Bars* to the *right* denote the efficiency of each template to drive faithful RNA initiation from the +1 site. Data represent the average values obtained in three to five independent transfections. Standard deviations are reported. Values were obtained by calculating first for each template, by PhosphorImager analyses, the +1/+6 transcript ratio, and subsequently by dividing such value by the level of +1/+6 transcripts driven by G3M.

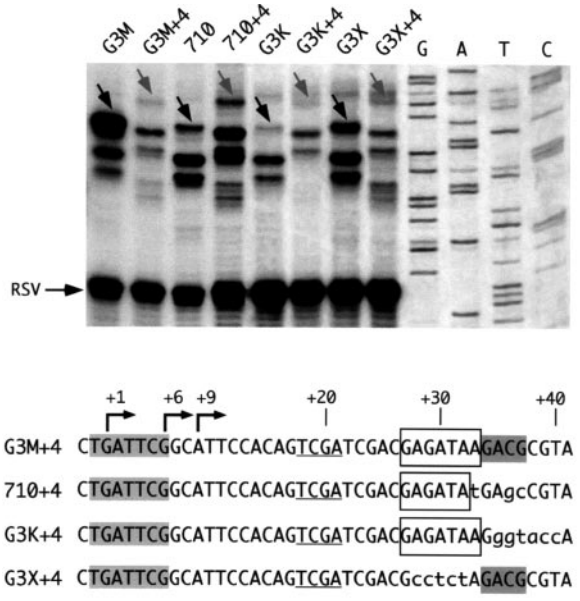

FIG. 5. **Space changes in the promoter region differently influence DPE and DE1 action.** Primer extension analysis of the transcripts directed in HEK 293 cells by the constructs indicated on *top* of the gel, and the G3M sequencing ladder, were obtained as described in the legends to the previous figures. *Black* and *gray arrows* within *lanes* mark +1 transcripts directed by parental and +4 derivative constructs, respectively. The G3M+4, 710+4, G3K+4, and G3X+4 promoter regions are shown at the *bottom*. Inr, DPE, and DE1 sequences are *highlighted* as in Fig. 2. The 4 bp inserted in each construct are *underlined*. Base changes altering residues found in G3M+4 are in *lowercase letters*.

levels of transcription initiation at +1. By contrast, substituting either residues +23 and +24 (construct 12A), or residues +27 (constructs 5A and 5B), +28 (construct 67B), and +29 (constructs 67E and 67K) had no major effect. However, when residues +28 and +29 were both changed, the levels of transcription initiation at +1 dropped 2–3-fold (constructs 67A, 67I, 67H, and 67G), except for the 67D. Thus, crucial residues are located both in the middle of DE1 and at the DE1/DPE boundary. Constructs carrying a mutated DPE motif could be broadly sorted in two main groups. The templates in which the first three DPE residues were preserved (constructs 71A and 71C, Fig. 4), resulted only ~2-fold less efficient than G3M, suggesting that DE1 can still efficiently cooperate with a partial DPE core. When only two adjacent residues of DPE were preserved, functional cooperation between DE1 and DPE was significantly reduced (constructs 710, 70A, 70T, and 701; Fig. 4),, yet three of these constructs, 710, 70A, and 70T, still directed initiation at +1 ~2-fold more efficiently than the DPE$^-$ G3K plasmid. Values are plausibly higher, as in all DPE mutants, the adenine at +29 was replaced by a thymine. The modification, which reduces only slightly the level of +1 transcripts (compare G3M and 67K constructs in Fig. 4), was introduced to enhance the effect of mutations hitting DPE without severely altering the DNA context.

On the whole, our data indicate that correct positioning of the transcriptional pol II machinery could be impaired at comparable levels by mutations affecting either DE1, DPE, or bases between the two motifs. Thus, DE1 and DPE appear to be part of a relatively large DNA region capable of multiple interactions with basal transcriptional factors, and it is therefore not surprising that most of the templates analyzed drive correct transcription initiation efficiently.

DPE, but Not DE1, Functions in a Strict Distance-dependent Fashion—There is a strict requirement for spacing between the Inr and DPE motifs, as an increase, or decrease, of a few nucleotides in the distance between the Inr and DPE causes a severe reduction in transcription. This suggests a specific and somewhat rigid interaction of TFIID with the Inr and DPE sequences (15–17). Interestingly, cooperation between Inr and DE1 is not strictly space-dependent. A pairwise comparison of the template activity of the four constructs G3M, G3X, G3K, and 710 with derivatives in which the distance separating Inr and downstream promoter sequences was increased by 4 bp, is reported in Fig. 5. In all constructs, the efficiency of *Inr*-dependent transcription was quantitated by comparing the levels of transcripts directed by the G3 Inr and the cotransfected RSVdel-CAT construct. Relatively to the parental G3M and G3X DPE$^+$ templates, faithful transcription initiation was reduced 12-and 4-fold in G3M+4 and G3X+4, respectively. By contrast, G3K and G3K+4, as the templates pair 710 and 710+4, all carrying DE1, directed the synthesis of faithfully initiated transcripts with the same efficiency (Fig. 5). An 8-bp increase in the distance between *Inr* and the downstream region abolished detectable initiation at +1 in all of the templates analyzed (data not shown).

TFIID-Promoter Interactions—To assess whether DE1$^+$ promoters interact with TFIID, the promoter regions of G3, G3M, G5M, G3K, and G3X plasmids were challenged with immunopurified holo-TFIID, and the formation of protein-promoter complexes assessed by electrophoretic mobility shift assays in agarose. In TFIID dose-response experiments, retarded complexes were detected with all the promoter regions assayed (Fig. 6, *panel A*). Each probe contains a GAL4 recognition sequence. The same amount of radiolabeled PCR product was separately incubated with a GAL4-NF-YA fusion protein (28), and retarded GAL4-NFYA/DNA complexes detected on 4.5% polyacrylamide gels (Fig. 6, *panel B*). Quantitative estimates were obtained by PhosphorImager analyses, and relative binding efficiencies were calculated by normalizing probe counts detected in retarded TFIID complexes to probe counts detected in GAL4 retarded complexes (Fig 6, *panel C*). By setting

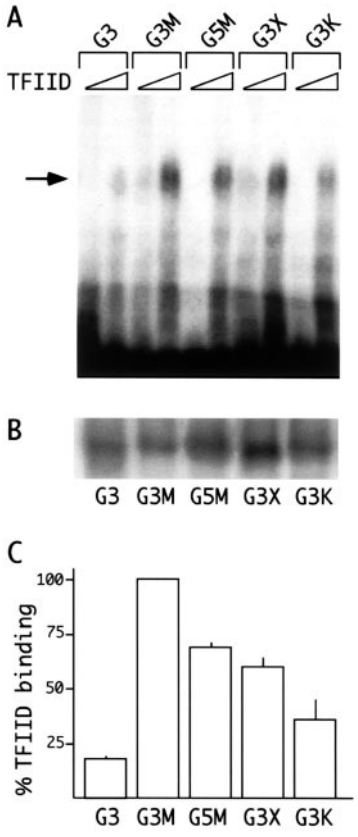

FIG. 6. **Representative agarose-EMSA analyses of TFIID-promoter complexes.** *A*, the ^{32}P-5′-end-labeled CAT II oligomer and the cold GT1 oligomer were used to amplify by PCR the −66/+61 region of the analyzed templates. Approximately 10,000 cpm of each purified PCR product was incubated with either 0.5 or 1.5 μl of an immunopurified human holo-TFIID fraction (see "Materials and Methods"), and samples were loaded onto a 1.2% agarose gel. Bands corresponding to TFIID-DNA retarded complexes are marked by an *arrow*. *B*, the same amount of radiolabeled PCR product was incubated with a GAL4-NF-YA fusion protein. Samples were loaded onto a 4.5% polyacrylamide gel, and GAL4-DNA retarded complexes are shown. *C*, relative TFIID binding efficiencies. Values result from the ratio of TFIID-DNA/GAL4-DNA complexes formed by each template at high TFIID input divided by the TFIID-DNA/GAL4-DNA complex ratio obtained with the G3M probe. Data represent the average values obtained in three to four independent experiments. Standard deviations are reported.

to 100% DNA/TFIID interactions detected with the DE1$^+$DPE$^+$G3M probe, we found that interactions of TFIID were 2-fold less efficient in the G3K template, in which the Inr is flanked by DE1 (36%), and in the G3X template, in which the Inr is flanked by DPE (51%). The efficiency of TFIID binding dropped ~5-fold in the absence of downstream promoter sequences (G3, 18% of binding). In accord to the transfection data shown in Fig. 5, DE1 and DPE are capable to interact with TFIID also in the absence of Inr sequences (G5M probe, 65% of binding).

DISCUSSION

Downstream promoter elements, often found in Inr-dependent promoters, function in part by increasing TFIID-promoter complex formation and/or stability through direct interactions with TAF_{II}s. A widely recognized downstream promoter element is DPE, a conserved motif found between residues +28 and +34 in many *Drosophila* transcriptional units. In the fruit fly, this DNA sequence is approximately as common as the TATA box (17). By contrast, inspection of the eukaryotic promoter data base (30) reveals that DPE-like sequences are rarely found, either at the described position or at alternative intragenic windows downstream of the site of RNA initiation, in mammalian promoters. Not surprisingly, DPE modules have been so far identified by functional analyses only in the human TATA-less promoters of the IRF-1 (15) and CD30 receptor (31) genes.

We thought it of interest to examine whether intragenic DNA sequences alternative to DPE, both in terms of sequence content and location relative to the RNA start site, could influence Inr-dependent transcription in a human cellular milieu. DE1 modules flank DPE in some *Drosophila* LINE promoters (16). In this work we showed that, in HEK 293 cells, the DE1 sequence GAGATAA spanning residues +23 to +29 stimulated transcription initiation from upstream Inr sequences nearly as efficiently as a core GACG DPE motif located at residues +30 to +33 (Figs. 2 and 4). Transcriptional enhancement significantly increased when the two motifs were adjacently located on the same template (Figs. 2 and 4). The stimulation, relatively mild on the strong G1 Inr, was magnified when the derivatives of the G3 plasmid were analyzed. In the absence of downstream activating modules, Inr sequences located at the same position of the G1 Inr sequences were not functional in G3 DNA, and transcription initiated preferentially from a secondary Inr (+6 and +9 transcripts, Fig. 1). By contrast, the primary G3 Inr was selectively stimulated by DPE as by DE1 (Fig. 2). Functional interactions between DPE and Inrs located at a specific distance are widely documented (12–16). The finding that DE1 mimics DPE in the activation of the same Inr is novel, and adds knowledge on the range of core promoter elements interacting with the pol II transcriptional machinery.

Results emerging from *in vivo* RNA analyses matched EMSA data, showing that DE1, as well as DPE, increased the stability of the TFIID-DNA complexes (Fig. 6). Qualitatively, the TFIID complexes were not grossly different among the templates used, suggesting that identical, or similar, TBP-containing complexes are involved. Quantitation of agarose-EMSAs revealed that retarded DNA-protein complexes resulting from the interaction with either Inr$^+$/DPE$^+$ or Inr$^+$/DE1$^+$ promoters were formed at comparable levels. TFIID-DNA complexes formed by promoter probes including both DE1 and DPE motifs were 2-fold more abundant (Fig. 6). These results are largely in agreement with the footprinting data previously reported on DPE-containing promoters (14, 15). On the whole, both transfection and biochemical data support the notion that DE1 is an intragenic signal analogous to DPE. Derivatives of the DE1$^+$/DPE$^+$ G3M promoter carrying alterations of DE1 directed faithfully initiated transcripts at least 2-fold more efficiently than the DE1$^-$/DPE$^+$ G3X promoter. Similar results were obtained by analyzing derivatives of the G3M promoter carrying mutated DPE motifs (Fig. 4). The observation that sequence contexts in which either DE1, DPE, or residues at the boundary of the two motifs are changed could sustain transcription with comparable efficiencies, supports the notion that multiple residues within the +23 to +33 region interact with TFIID. A key feature of DPE-driven core promoters is a precise spacing between the Inr and DPE. Noteworthy, functional Inr/DE1 interactions are not as strictly space-dependent as Inr/DPE interactions (Fig. 5). This observation leads us to hypothesize that DE1 and DPE, which cannot be exposed on the same side of the DNA double helix, may interact with different surfaces of TFIID, and possibly contact different TAF_{II}s. DPE is bound by TAF_{II}60-TAF_{II}40 heterotetramers (15). DE1 may stimulate transcription by contacting either TAF_{II}250 orTAF_{II}150, the two TFIID components involved in recognition of the Inr and sequences further downstream (8, 20), although at present a role in promoter recognition for some of the other TAF_{II}s cannot be excluded.

TFIID binds to core promoters through interactions that are

apparently multiple, in that the TATA, Inr, and DPE elements have all been clearly shown to be associated with distinct subunits of the complex (TBP, $TAF_{II}150$-$TAF_{II}250$, and $TAF_{II}60$-$TAF_{II}40$, respectively). This combination of elements serves to maximize TFIID stability on the promoter, thereby contributing to promoter strength (see Ref. 22). It is not, therefore, surprising that both DE1 and DPE may cooperate with Inr to sustain transcription, although the precise rules allowing one element to work in some promoters in the apparent absence of additional contacts are poorly understood. One possibility is that additional factors potentially binding to TFIID, such as TFIIA and NC2, will help in fine-tuning the interactions with core promoter elements, both in a positive and negative way. Intragenic core promoter elements distinct from DPEs have been described in a few promoters. In the human $TATA^+$ megalin/low density lipoprotein receptor-related protein 2 gene, promoter sequences located between positions +5 and +11 (5′-TTTTGGC-3′) interact with TFIID. Downstream contacts do not significantly affect the overall affinity of TFIID binding, but induce dramatic qualitative changes in TFIID interactions in the lipoprotein receptor-related protein 2 TATA box region (21). The human $TATA^+Inr^+$ β-globin promoter contains a large downstream region interacting with TFIID called DCE. Functional DCE subelements map at positions +13/+15, +22/+24, and +31/+33 (see Ref. 22). In the *Drosophila* hsp70 promoter, four regions interact with TFIID: the TATA element, the initiator, and two regions located ~18 and 28 nucleotides downstream of the transcription start site (19). In transgenic flies, Inr and downstream sequences serve overlapping functions, making rather modest contributions to the level of expression of the hsp70 promoter (32). The contributions of individual core sequences could have significant physiological impact in other promoters, and mutations in the β-globin gene DCE subelements II and III are the basis for two kinds of human thalassemia (22). The finding that sequences capable to interact with TFIID found at the same gene coordinates may differently contribute to promoter strength *in vivo* illustrates the difficulties in predicting the functional architecture of core promoters.

Plausibly, DE1-like sequences are not restricted to LINE promoters. This is supported both by statistical and biochemical analyses, indicating that a G nucleotide located 4 bp upstream of the DPE core contributes to transcription from DPE-containing promoters (17). Interestingly, in the G3M promoter, the central G of DE1 is at 4 bp distance from DPE (residue +25, Fig. 2), and is important for DE1 activity (Fig. 4). Notably, a DE1-like motif (5′-GAGGCAA-3′) immediately flanks DPE in the human IRF-1 gene and may account for the residual activity of the IRF-1 promoter upon removal of DPE (15). Finally, a purine-rich sequence partly resembling DE1 (5′-GAGACG-3′) is located at residues +23 to +28 in the middle of the downstream region of the human *gfa* (glial fibrillary acid) promoter, also interacting with TFIID (18). Presumably, DE1 sequences are common to many promoters, but are overlooked because of their poor homology. The consensus resulting from the alignments of three $DE1^+$ *Drosophila* LINE promoters is relatively loose (5′-GRG(A/T)(G/T)AA-3′; see Ref. 16), and different sequences may have DE1 activity, as emerging from the analysis of mutagenized templates in Fig. 4. Sequence flexibility has been similarly observed for DPEs, because the range of sequences that can function as a DPE extend well beyond the GA/TCG motif (17). The analysis of randomized promoter libraries may help to determine the range of functional DE1 sequences and derive position weight matrices used to predict the occurrence of analogous modules in natural promoters as done for TATA and Inr elements (2, 6).

Future analyses may reveal whether DE1 sequences are predominantly found in isolation, or associated to DPE motifs. Transcriptional enhancers that are specific for promoters that contain either DPE or TATA box elements have been elegantly identified by P-mediated transformation analyses in *Drosophila* (33), and it has been shown that the transcriptional repressor NC2 activates DPE-driven promoters and represses TATA-driven promoters *in vitro* (34). In light of these findings, it would be of interest to ascertain whether $DE1^+DPE^+$ and DPE^+ promoters may functionally differ in some of these properties *in vivo*.

Acknowledgment—We thank Valerio Orlando for suggestions and critical reading of the manuscript.

REFERENCES

1. Roeder, R. G. (1996) *Trends Biochem. Sci.* **21,** 327–335
2. Bucher, P. (1990) *J. Mol. Biol.* **212,** 563–578
3. Zawel, L., and Reinberg, D. (1995) *Annu. Rev. Biochem.* **64,** 533–561
4. Lagrange, T., Kapanidis, A. N., Tang, H., Reinberg, D., and Ebright, R. H. (1998) *Genes Dev.* **12,** 34–44
5. Smale, S. T. (1997) *Biochim. Biophys. Acta* **1351,** 73–88
6. Kraus, R. J., Murray, E. E., Wiley, S. R., Zink, N. M., Loritz, K., Gelembiuk, G. W., and Mertz, J. E. (1996) *Nucleic Acids Res.* **24,** 1531–1539
7. Kaufmann, J., and Smale S. T. (1994) *Genes Dev.* **8,** 821–829
8. Chalkley, G. E., and Verrijzer, C. P. (1999) *EMBO J.* **18,** 4835–4845
9. Kollmar, R., Sukow, K. A., Sponagle, S. K., and Farnham. P. J. (1992) *J. Biol. Chem.* **269,** 2252–2257
10. Colgan, J., and Manley, J. L. (1995) *Proc. Natl. Acad. Sci. U. S. A.* **92,** 1955–1959
11. Emami, K. H., Navarre, W. W., and Smale, S. T. (1995) *Mol. Cell. Biol.* **15,** 5906–5916
12. Minchiotti, G., and Di Nocera, P. P. (1991) *Mol. Cell. Biol.* **11,** 5171 5180
13. Contursi, C., Minchiotti, G., and Di Nocera, P. P. (1995) *J. Biol. Chem.* **270,** 26570–26576
14. Burke, T. W., and Kadonaga, J. T. (1996) *Genes Dev.* **10,** 711–724
15. Burke, T. W., and Kadonaga, J. T. (1997) *Genes Dev.* **11,** 3020–3031
16. Minchiotti, G., Contursi, C., and Di Nocera, P. P. (1997) *J. Mol. Biol.* **267,** 37–46
17. Kutach, A. K., and Kadonaga, J. T. (2000) *Mol. Cell. Biol.* **20,** 4754–4764
18. Nakatani, Y., Horikoshi, M., Brenner, M., Yamamoto, T., Besnard, F., Roeder, R. G., and Freese, E. (1990) *Nature* **348,** 86–88
19. Purnell, B. A., Emanuel, P. A., and Gilmour, D. S. (1994) *Genes Dev.* **8,** 830–842
20. Verrijzer, C. P., Yokomori, K., Chen, J. L., and Tjian, R. (1994) *Science* **264,** 933–941
21. Knutson, A., Castaño, E., Oelgeschläger, T., Roeder, R. G., and Westin, G. (2000) *J. Biol. Chem.* **275,** 14190–14197
22. Lewis, B. A., Kim, T. K., and Orkin, S. H. (2000) *Proc. Natl. Acad. Sci. U. S. A.* **97,** 7172–7177
23. Chomczynski, P., and Sacchi, N. (1987) *Anal. Biochem.* **162,** 156–159
24. Grimaldi, G., and Di Nocera, P. P. (1988) *Proc. Natl. Acad. Sci. U. S. A.* **85,** 5502–5506
25. Jacq, X., Brou, C., Lutz, Y., Davidson, I., Chambon, P., and Tora, L. (1994) *Cell* **79,** 107–117
26. Wieczorek, E., Brand, M., Jacq, X., and Tora, L. (1998) *Nature* **393,** 187–191
27. Frontini, M., Imbriano, C., di Silvio, A., Bell, B., Bogni, A., Romier, C., Moras, D., Tora, L., Davidson, I., and Mantovani, R. (2002) *J. Biol. Chem.* **277,** 5841–5848
28. Liberati, C., di Silvio, A., Ottolenghi, S., and Mantovani, R. (1999) *J. Mol. Biol.* **285,** 1441–1455
29. Minchiotti, G., and Di Nocera, P. P. (1997) *FEBS Lett.* **411,** 189–194
30. Perier, R. C., Junier, T., Bonnard, C., and Bucher, P. (1999) *Nucleic Acids Res.* **27,** 307–309
31. Croager, E. J., Gout, A. M., and Abraham, L. J. (2000) *Am. J. Pathol.* **156,** 1723–1731
32. Wu, C. H., Madabusi, L., Nishioka, H., Emanuel, P., Sypes, M., Arkhipova, I., and Gilmour, D. S. (2001) *Mol. Cell. Biol.* **21,** 1593–1602
33. Butler, J. E., and Kadonaga, J. T. (2001) *Genes Dev.* **15,** 2515–2519
34. Willy, P. J., Kobayashi, R., and Kadonaga, J. T. (2000) *Science* **290,** 982–985

GENES: STRUCTURE AND REGULATION:

A Novel Intragenic Sequence Enhances Initiator-dependent Transcription in Human Embryonic Kidney 293 Cells

Chiara Abrescia, Eliana De Gregorio, Mattia Frontini, Roberto Mantovani and Pierpaolo Di Nocera

J. Biol. Chem. 2002, 277:19594-19599.
doi: 10.1074/jbc.M201193200 originally published online March 26, 2002